Accounting-Based Distress Prediction Under Target Leakage: A Three-Estimand Framework with Evidence from Colombian Private Firms

Future Accounting Distress in Colombian Private Firms:

A Parsimonious Two-Ratio Score with Three-Estimand Framework

Jhon Alexander Jiménez Triviño, Ph.D.

Investment Banking and Corporate Valuation
Colombia

This version: May 2026

HIGHLIGHTS

▪ We propose a three-estimand framework for accounting-based distress prediction distinguishing contemporaneous classification, forward-looking prediction, and transition prediction; the empirical gap among them quantifies the contribution of target leakage to apparent in-sample performance.

▪ Sample of 192,458 firm-year observations from Supersociedades Colombia, 2018-2024, spanning 37,653 unique formal corporates across 87 industry divisions.

▪ The three estimands yield AUCs of 0.955 (contemporaneous, with definitional leakage), 0.826 (forward-looking, 95% CI [0.819, 0.829]) and 0.654 (transition, 95% CI [0.643, 0.661]); the 13-30 percentage-point gap is interpreted as the leakage component, not as competing model performance.

▪ Walk-forward out-of-time validation across 2019-2023 yields AUCs of 0.78-0.85, including the COVID-19 shock year (test 2020 AUC = 0.84).

▪ On a separate historical Colombian panel where all Altman (1968, 1995) inputs are available, the local logit improves AUC by 15-19 percentage points over correctly-oriented Altman benchmarks (paired-bootstrap $p < 0.005$).

▪ An external cross-reference against 3,575 legal insolvency events from the Supersociedades public registry (2022-2025) - events not used for estimation - finds that 80.5% of subsequent judicial liquidations had predicted PD >= 10% one year prior; this is asymmetric evidence conditional on observed events.

▪ The framework is portable to any predictive setting where the predictor set and the outcome share an accounting identity. The Colombian point estimates are not recommended for entities in CIIU divisions K64 and K66 (monetary intermediation and financial auxiliaries), where balance-sheet structures saturate the capitalization ratio.

ABSTRACT

Accounting-based distress prediction is methodologically delicate when predictors and outcomes share accounting variables: the resulting target leakage inflates apparent in-sample performance and obscures the genuine predictive content of the model. We propose a three-estimand framework - contemporaneous classification, forward-looking prediction, and transition prediction - that quantifies and partially neutralizes this dependence. Using 192,458 firm-year observations from the Colombian regulator Supersociedades for 2018-2024, we estimate a parsimonious two-ratio logit combining operating profitability (EBIT/Total Assets) and capitalization (Equity/Total Assets) on 37,653 unique formal corporates. The three estimands yield AUCs of 0.955 (contemporaneous, with definitional leakage), 0.826 (forward-looking, 95% CI [0.819, 0.829]) and 0.654 (transition, 95% CI [0.643, 0.661]); the 13-30 percentage-point gap quantifies the magnitude of the leakage component. Walk-forward out-of-time validation across 2019-2023 yields AUCs of 0.78-0.85, including the COVID-19 year. On a separate historical Colombian panel where all Altman (1968, 1995) inputs are available, the local logit improves AUC by 15-19 percentage points over correctly-oriented Altman benchmarks (paired-bootstrap $p < 0.005$). An external cross-reference against 3,575 legal insolvency events from the Supersociedades public registry (2022-2025) - events not used for estimation - finds that 80.5% of subsequent judicial liquidations had predicted PD >= 10% one year before opening; this is asymmetric evidence conditional on observed events and does not quantify false positives across the population. The framework is the primary contribution; the Colombian estimates illustrate its application. The score is not recommended for entities in CIIU divisions K64 and K66, where balance-sheet structures saturate the capitalization ratio.

Keywords: Financial distress; target leakage; forward-looking validation; transition prediction; Colombia; emerging markets; logit; survivorship bias; calibration.

JEL Classification: G33 (Bankruptcy; Liquidation); G32 (Capital Structure); M41 (Accounting); C52 (Model Evaluation).

I. Introduction

Problem. Predicting financial distress in private firms in emerging markets is empirically difficult. Market prices that would aggregate forward-looking expectations are unavailable; accounting data are subject to reclassifications, revaluations, and varying audit quality; the inferential target itself is often ambiguously defined. Colombia exemplifies the setting: fewer than one hundred firms are listed on the Bolsa de Valores de Colombia, while approximately 30,000-37,000 corporates file annual financial statements to the Superintendencia de Sociedades (Supersociedades) under regulatory thresholds. The bulk of credit risk in the Colombian formal corporate sector is therefore borne by private firms whose distress must be inferred from accounting fundamentals alone.

Gap in the literature. A substantial accounting-based distress literature exists, originating with Beaver (1966) and Altman (1968) and extended via logit (Ohlson, 1980), hazard models (Shumway, 2001; Campbell, Hilscher, and Szilagyi, 2008), and emerging-market adaptations (Pereiro, 2001, 2010). What this literature has not systematically addressed - and what we operationalize here - is the inferential consequence of an endemic feature: when the outcome of interest (distress) is defined using accounting variables that also enter as predictors, the resulting target leakage inflates apparent performance by an amount that is not visible in any single regression but emerges as a gap across estimands that differ in their temporal and conditional structure.

Target leakage formally. Kaufman, Rosset, and Perlich (2012) define target leakage as the inadvertent inclusion of outcome information in predictors. In accounting-based distress prediction this dependence is not avoidable by feature engineering when distress is defined as, for example, severe operating losses ($R1 <$ threshold) or negative equity ($R4 < 0$) - the same ratios serve as predictors. Forward-looking validation (predicting t+1 from t) breaks the temporal coincidence but persistence in the distress state preserves a channel of mechanical dependence. The cleanest test is transition prediction: conditional on a firm being healthy at t, predict deterioration at t+1. The empirical gap between contemporaneous and transition AUCs measures the magnitude of the leakage.

Data and method. We use the Supersociedades Estados Financieros database for 2018-2024 (post-IFRS), yielding 192,458 firm-year observations from 37,653 unique firms across 87 two-digit CIIU industry divisions. We estimate a parsimonious two-ratio logit (R1 = EBIT/Total Assets, R4 = Equity/Total Assets) under three outcomes: contemporaneous classification at t, forward-looking prediction at t+1, and transition prediction conditional on being healthy at t. Predictors are winsorized at the 1st and 99th percentiles; standard errors are clustered by firm; walk-forward out-of-time validation re-estimates the model annually with winsorization thresholds computed only on the training window.

Main results. The three estimands yield AUC values of 0.955, 0.826 (95% CI [0.819, 0.829]) and 0.654 (95% CI [0.643, 0.661]), respectively; the 13-30 percentage-point gap is interpreted as the magnitude of definitional dependence rather than as competing model performance. Walk-forward AUC across 2019-2023 ranges from 0.78 to 0.85, including the COVID-19 year. On a separate historical Colombian panel (2006-2018) where all Altman (1968, 1995) inputs are available, the local two-ratio logit improves AUC by 15-19 percentage points over correctly-oriented Altman benchmarks (paired-bootstrap $p < 0.005$). A third candidate predictor (operating return on equity) contributes essentially no additional forward-looking AUC. A Cox proportional-hazards specification with time-varying R1 and R4 is reported as duration-based robustness; its apparent AUC advantage over the logit is shown to reflect local recalibration rather than a structural property of the duration form. An external cross-reference against the Supersociedades public registry of 3,575 legal insolvency events (2022-2025) - events not used in estimation - finds that 80.5% of subsequent judicial liquidations had predicted PD >= 10% one year prior; we discuss the asymmetric nature of this evidence in Section V.J.

Contribution. Our primary contribution is methodological: we propose and apply a three-estimand framework that distinguishes contemporaneous classification, forward-looking prediction, and transition prediction, and we document the magnitude of the inferential gap among them in a setting where target leakage is endemic. The framework is portable to any predictive problem in which the predictor set and the outcome share an accounting identity. The Colombian point estimates illustrate the framework rather than constituting the contribution. Secondary contributions are the apples-to-apples Altman benchmarking on a separate historical panel where all Altman inputs are available, a formal sensitivity analysis of attrition-induced selection, and the external cross-reference against legal insolvency events (Section V.J).

Scope. We do not claim that any of our outcomes coincides with legal insolvency. The score is calibrated on accounting distress in formal corporates above Supersociedades reporting thresholds; it is not recommended for entities classified in CIIU divisions K64 (monetary intermediation) and K66 (financial auxiliaries), where balance-sheet structures saturate the capitalization ratio.

Organization. Section II reviews the relevant literature. Section III describes the data, sample construction, and panel attrition. Section IV presents the three outcome definitions and the methodological framework. Section V reports leakage diagnostics, the parsimonious specification with its operational formula, walk-forward validation, calibration and population stability, Altman benchmarking, sectoral robustness, a Cox specification as duration-based robustness, and the external cross-reference against legal insolvency events. Section VI discusses implications and limitations. Section VII concludes.

II. Related Literature

A. Foundational Distress Prediction Models

The accounting-based distress prediction literature originates with Beaver (1966) and Altman (1968). Altman applied Multivariate Discriminant Analysis (MDA) to 66 U.S. manufacturing firms, yielding the celebrated Z-score: a linear combination of Working Capital/TA, Retained Earnings/TA, EBIT/TA, Market Value of Equity/Total Liabilities, and Sales/TA. Altman's Z' (1983) and Z" (1995) variants modified the specification for private and emerging-market firms by replacing market-based leverage with book-value leverage and dropping the sales-to-assets ratio.

Ohlson (1980) replaced MDA with logistic regression, arguing that the latter does not require multivariate normality or equal covariance matrices. Logit became the standard for academic work, while MDA persists in industry due to its interpretability. Shumway (2001) introduced discrete-time hazard models that exploit the panel structure of firm data and incorporate time-varying covariates. Campbell, Hilscher, and Szilagyi (2008) further refined this framework, achieving AUC values around 0.88 on U.S. data using a combination of accounting and market-based predictors.

B. Distress Prediction in Emerging Markets

Pereiro (2001, 2010) emphasizes that local calibration is essential for emerging markets, where fundamentals differ from developed-market settings. For Colombia, Sarmiento-Sabogal and

Sadeghi (2014, 2015) and Sarmiento-Sabogal et al. (2021) develop methods for estimating the cost of equity for non-listed firms using accounting fundamentals. Caicedo-Cerezo (2006) provides one of the first systematic distress prediction studies for Colombia, but with a substantially smaller sample than ours. Bai and Green (2020) document that country and industry factors operate as independent drivers of stock returns in partially integrated emerging markets, with implications for the validity of imported CAPM-based cost-of-equity estimates.

C. Target Leakage and Forward-Looking Validation

A critical methodological issue in applied predictive modeling is target leakage: the unintended inclusion of information about the outcome in the predictors. Kaufman et al. (2012) describe leakage as among the most common errors in real-world predictive modeling. In distress prediction, leakage commonly arises when the outcome is defined using accounting variables that also appear as predictors. The remedy is forward-looking validation: predicting outcome at t+1 using predictors observed at t (Hyndman and Athanasopoulos, 2018). Walk-forward cross-validation provides an additional safeguard against information leakage from the future to the past.

III. Data and Sample

A. Source and Variable Construction

The Supersociedades Estados Financieros database is the regulatory repository for Colombian formal corporate entities exceeding statutory reporting thresholds. We restrict our analysis to the post-IFRS period (2018-2024) for accounting consistency. We require non-missing values for total assets, total liabilities, equity, operating income, operating revenues, and gross profit, and we drop observations with non-positive total assets. We winsorize all financial ratios at the 1st and 99th percentiles to mitigate extreme outliers, following Adams et al. (2019). For walk-forward validation, winsorization thresholds are estimated using only the training sample (years through t-1) and then applied to the test year, preventing any look-ahead bias in out-of-time evaluation.

Three accounting ratios form the basis of our analysis:

$$R1 = \textit{Operating Income} / \textit{Total Assets} \qquad (1)$$

$$R4 = \textit{Total Equity} / \textit{Total Assets} \qquad (2)$$

$$ROE\_op = \textit{Operating Income} / \textit{Total Equity} \qquad (3)$$

R1 captures operating profitability scaled by asset size; R4 measures capitalization (the inverse of leverage); ROE_op is mathematically equivalent to R1/R4 and captures capital efficiency. ROE_op is undefined when equity equals zero and can take extreme values when equity is small or negative; we mitigate this through winsorization. Section V.B critically evaluates the contribution of ROE_op and ultimately recommends the parsimonious R1+R4 specification.

B. Three Distress Outcome Definitions

Outcome A (contemporaneous classification, t):

$$DISTRESS\_t = 1 \textit{ if } (E\_t < 0) \textit{ OR } (TL\_t > TA\_t) \textit{ OR } (R1\_t < -0.10) \qquad (4)$$

Outcome B (forward-looking prediction, t+1):

$$DISTRESS\_\{i,t+1\} = \textit{same conditions evaluated at year } t+1, \textit{ predictors at } t \qquad (5)$$

Outcome C (transition prediction):

$$TRANSITION\_\{i,t+1\} = DISTRESS\_\{i,t+1\} \mid DISTRESS\_\{i,t\} = 0 \qquad (6)$$

Outcome A is acknowledged as a financial health classifier with definitional dependence on R1 and R4 (Section V.A documents the magnitude). Outcome B is the genuine one-period forecast; the definitional dependence is broken by temporal separation, although autocorrelation in distress states preserves a persistence channel. Outcome C is the most demanding metric: it captures only firms transitioning into distress from previously healthy states.

We make no claim that any of these outcomes coincides with legal insolvency. Supersociedades does not flag legal events of liquidation, reorganization, or dissolution in the standard data extract; we therefore restrict our outcomes to accounting-based definitions and acknowledge this limitation in Section VI.B.

C. Sample Construction, Reconciliation, and Attrition

The full filtered analytical sample comprises 192,458 firm-year observations spanning 37,653 unique firms across 87 two-digit CIIU industry divisions. The raw extract contains 37,783 unique firms before analytical filters. We construct the forward-looking panel by linking each observation in year t to the same firm's outcome in year t+1. Firms not observed in year t+1 are excluded from the forward-looking panel (151,710 firm-year observations). The transition panel is further restricted to firms not currently in distress (142,214 firm-year observations).

Table I provides a sample-flow reconciliation that documents the relationship between filing counts at each step of construction.

Table I

Sample Reconciliation (firm-year observations)

Sample reconciliation across analytical panels. Each row reports total firm-year observations after a step. The forward-looking panel includes firm-years (i, t) with both year-t financial statements and year-(t+1) outcome observable. The transition panel further restricts to firms with DISTRESS_t = 0, isolating new deterioration.

Step	Firm-year obs	Notes
Raw 2018-2024 (post filters)	192,458	37,653 final analytical firms × 87 CIIU divisions
Forward-looking panel (Outcome B)	151,710	(i, t) with t+1 outcome observable
Transition panel (Outcome C)	142,214	Subset with DISTRESS_t = 0

No observable t+1 outcome	40,748	Final-year observations and firms exiting before t+1

Using strict NIT-year linking, the forward-looking panel contains 151,710 firm-year observations, consistent with the consecutive-year intersections reported in Table II. Observations without an observable t+1 outcome consist of final-year observations and firms that exit the reporting panel before the following year.

The implicit conditioning on firms observed in t+1 introduces potential survivorship bias. Table II documents annual attrition: on average, 7.7% of firms observed in year t are not observed in t+1, with an annual range of 5.5% to 10.3%. Critically, the average distress rate among firms exiting the panel (23.0%) is approximately 3.7 times the rate among firms persisting (6.3%). This asymmetry has two interpretations. First, some attrited firms may have exited because of legal insolvency or liquidation, which our data cannot directly observe. Second, firms in financial distress may shrink below regulatory reporting thresholds and exit the panel without legal default. The direction of the resulting bias in our AUC estimates is theoretically ambiguous: excluding distressed attriters could either inflate AUC (by removing hard-to-classify cases) or deflate it (by removing easy-to-classify cases). We therefore present our forward-looking estimates as conditional on firms surviving in the reporting panel through t+1, and we interpret AUC values accordingly.

Table II

Annual Panel Attrition and Distress Asymmetry

Annual attrition statistics for unique firms. N_t reports firms observed in year t; Persisting reports firms also observed in t+1; Attrited reports firms in t but not in t+1. The two rightmost columns report the distress rate at year t separately for attrited and persisting firms.

Year (t)	N in t	Persisting	Attrited	Attrition rate	Distress | attrited	Distress | persisting

2018	19,263	17,941	1,322	6.9%	21.9%	7.4%
2019	25,416	23,183	2,233	8.8%	19.3%	5.9%
2020	28,882	26,762	2,120	7.3%	24.5%	6.8%
2021	28,954	26,866	2,088	7.2%	20.6%	5.9%
2022	30,843	29,134	1,709	5.5%	31.1%	6.2%
2023	31,019	27,824	3,195	10.3%	20.8%	5.8%
Avg	—	—	—	7.7%	23.0%	6.3%

Table III provides summary statistics for the predictors and outcomes.

Table III

Summary Statistics

Descriptive statistics for the variables used in the analysis. R1 = Operating Income / Total Assets; R4 = Equity / Total Assets; ROE_op = Operating Income / Equity. All ratios winsorized at 1st and 99th percentiles.

Variable	Mean	Std. Dev.	25th	75th	N
R1 (EBIT/TA)	0.046	0.114	-0.006	0.107	192,458
R4 (E/TA)	0.398	0.299	0.197	0.625	192,458
ROE_op	0.122	0.461	-0.026	0.295	192,458
DISTRESS_t (Outcome A)	0.0745	0.263	—	—	192,458
DISTRESS_{t+1} (Outcome B)	0.0699	0.257	—	—	151,710
TRANSITION_{t+1} (Outcome C)	0.0323	0.178	—	—	142,214

IV. Methodology

A. Main Specification: Logit

Following Ohlson (1980), our main model is logistic regression. We estimate two specifications: the parsimonious recommended model M1 and a three-variable model M2 retained for comparison with the prior literature.

$$M1: P(DISTRESS\_\{i,t+1\} = 1 \mid X\_\{i,t\}) = \Lambda(\alpha + \beta_1 \cdot R1\_\{i,t\} + \beta_2 \cdot R4\_\{i,t\}) \quad (7)$$

$$M2: P(DISTRESS\_\{i,t+1\} = 1 \mid X\_\{i,t\}) = \Lambda(\alpha + \beta_1 \cdot R1\_\{i,t\} + \beta_2 \cdot R4\_\{i,t\} + \beta_3 \cdot ROE\_op\_\{i,t\}) \quad (8)$$

where Λ(·) denotes the logistic CDF. Standard errors are clustered by firm identifier (NIT) to account for the panel structure of the data.

B. Pooled Logit with Year and Sector Fixed Effects

We additionally estimate a pooled logit with year and sector fixed effects (M3). Earlier drafts characterized this specification as 'hazard-style' following the discrete-time framework of Shumway (2001); given the genuine Cox proportional hazards specification reported in Section V.I, we adopt the more precise terminology 'pooled logit with year and sector fixed effects.' Year fixed effects capture macroeconomic shocks common to all firms in a given year, and sector fixed effects capture time-invariant industry differences.

$$M3: P(DISTRESS\_\{i,t+1\} = 1) = \Lambda(\alpha + \beta_1 \cdot R1\_\{i,t\} + \beta_2 \cdot R4\_\{i,t\} + \gamma\_y + \delta\_s) \quad (9)$$

We include year dummies for 2019-2023 (with 2018 as the omitted base) and sector dummies for the top 30 two-digit CIIU divisions (with 'OTHER' as the omitted base). Standard errors clustered by NIT.

C. Validation Framework

We employ four validation strategies. First, we report contemporaneous in-sample performance (Outcome A) as a benchmark. Second, we report forward-looking in-sample performance (Outcome B). Third, we report transition in-sample performance (Outcome C). Fourth, we conduct walk-forward out-of-time validation: for each test year t, we train the model on observations from 2018 through t-1 and evaluate on year t with outcome at t+1.

Performance metrics include the Area Under the ROC Curve (AUC) with 95% bootstrap confidence intervals (500 resamples), the Brier Score, the Brier Skill Score (BSS), the Expected and Maximum Calibration Errors (ECE and MCE), and decile-level calibration tables. We assess differences in AUC between local and Altman benchmarks via paired bootstrap, reporting bootstrap p-values for ΔAUC. Population stability is assessed via the Population Stability Index (PSI) computed year-over-year against a 2018 baseline. A note on reported AUCs across analyses: the main forward-looking AUC of 0.826 is estimated on the strict forward-looking panel of 151,710 firm-year observations (firms with both year-t financial statements and observable year-(t+1) outcome). Slight differences in AUC across alternative analyses (0.810 in the duration sub-panel of Section V.I; 0.821 in the Altman recalibration sub-panel of Section V.G, restricted to 161,253 firm-year observations with all Altman inputs available; 0.845 in the historical 2006-2018 comparison panel of Section V.G, where pre-IFRS accounting yields sharper discrimination) reflect subsample definitions adopted for each specific analysis; the 0.826 figure is the primary reported metric. Bootstrap-based AUC differences are tested via paired bootstrap (DeLong et al., 1988).

V. Empirical Results

A. Leakage Diagnostics

We begin by quantifying definitional leakage in Outcome A. The distress indicator combines three conditions: negative equity (4.31% of observations), liabilities exceeding assets (4.52%), and severe operating losses (R1 < -0.10, 4.62%). The first two conditions are mathematically equivalent under the accounting identity TA = TL + E and indeed coincide in 99.7% of observations (the 0.3% discrepancy reflects rounding in regulatory filings and reclassification adjustments that occasionally break the strict accounting identity in audited reports). The third condition is a direct function of predictor R1.

Approximately 39% of distressed observations are classified solely by the third condition (R1 < -0.10), and an additional 35% solely by the equity-based conditions. The two condition sets jointly classify the remaining 26%. Moreover, R4 < 0 implies DISTRESS = 1 with probability one. These constitute direct mechanical relationships between predictors and outcome, not statistical patterns.

The empirical impact of this leakage is large. We estimate the same logit specification with three different outcomes, holding predictors constant. With Outcome A (contemporaneous), AUC = 0.9549. With Outcome B (forward-looking), AUC = 0.8257—a decline of 12.9 percentage points. With Outcome C (transition), AUC = 0.6543—a further decline of 17.1 percentage points. We interpret the contemporaneous AUC as reflecting both genuine predictor information and definitional dependence; the forward-looking AUC as reflecting prediction with persistence; and the transition AUC as the cleanest test of new-deterioration prediction.

B. Parsimonious Two-Ratio Specification and Operational Formula

Table IV reports coefficient estimates for both the three-variable specification M2 (with ROE_op) and the parsimonious two-variable specification M1 (without ROE_op), under both Outcome B (forward-looking) and Outcome C (transition).

Table IV

Logit Coefficients: Forward-Looking and Transition (Honest Significance)

*Logit estimates of distress probability. Panel A (forward) uses Outcome B, N = 151,710. Panel B (transition) uses Outcome C, N = 142,214. Standard errors clustered by NIT (firm). z-statistics in parentheses. Significance: *** p<0.01, ** p<0.05, * p<0.10.*

Variable	M2 Forward (3 vars)	M1 Forward (R1+R4)	M2 Transition	M1 Transition (R1+R4)
const	-1.1895*** | (-52.61)	-1.2011*** | (-54.45)	-2.4059*** | (-67.47)	-2.4172*** | (-68.90)
R1	-6.5162*** | (-	-6.7120*** | (-	-3.6769*** | (-	-3.9491*** | (-

	37.99)	43.28)	9.23)	13.02)
R4	-3.4455*** | (-56.09)	-3.4259*** | (-56.48)	-1.5166*** | (-22.26)	-1.4902*** | (-23.01)
ROE_op	-0.0889*** | (-3.49)	—	-0.0885 | (-1.06)	—
N	151,710	151,710	142,214	142,214
AUC	0.8256	0.8257	0.6514	0.6543

Three observations follow from Table IV. First, R1 and R4 are strongly significant ($p < 0.001$) in all four specifications, with the expected negative sign: higher operating profitability and higher capitalization reduce predicted distress probability. Second, ROE_op is significant in the forward-looking specification ($p < 0.001$) but not in the transition specification ($\beta = -0.0885$, SE = 0.0832, $z = -1.06$, $p = 0.29$). Third, dropping ROE_op leaves AUC essentially unchanged in both specifications: forward-looking AUC moves from 0.8256 to 0.8257, and transition AUC moves from 0.6514 to 0.6543.

A formal likelihood-ratio test of M1 against M2 in the forward-looking sample yields LR = 18.92 (df = 1, $p = 0.000014$), apparently rejecting M1 in favor of M2. However, this rejection is a large-N artifact: with 151,710 observations, even economically negligible differences become statistically significant. The ΔAUC is effectively zero, and the ROE_op coefficient is insignificant in transition. We conclude that ROE_op contributes no meaningful predictive information beyond what R1 and R4 jointly provide. The recommended forecasting specification is therefore the parsimonious M1.

B.1. Operational Scoring Formula

The recommended forecasting score is operationalized as follows. For a firm with operating profitability R1 = EBIT_t / TA_t and capitalization R4 = Equity_t / TA_t, the predicted probability of accounting distress at t+1 is:

$$z = -1.2011 + (-6.7120)\cdot R1 + (-3.4259)\cdot R4 \qquad (10)$$

$$P(DISTRESS\_\{t+1\} = 1) = 1 / (1 + \exp(-z)) \qquad (11)$$

Both slope coefficients and the intercept are highly significant under cluster-robust standard errors by firm. The estimated AUC is 0.8257 (95% bootstrap CI [0.8189, 0.8294]). For applications restricted to firms not currently in distress, the analogous transition-specification formula and worked numerical examples are provided in Appendix A.

C. Statistical Tests of the Final Model

Table V compiles the formal statistical tests on the final M1 specification. Bootstrap 95% confidence intervals for AUC are computed with 500 resamples; the narrow interval ([0.819, 0.829]) confirms precise estimation. The Wald joint test of β_R1 = β_R4 = 0 yields $\chi^2 = 5{,}894$ (df = 2, $p < 0.001$), strongly rejecting the null. We report Hosmer-Lemeshow with the standard caveat that the test rejects calibration at large N even for well-calibrated models; ECE-based metrics are more reliable in our setting (Hosmer and Lemeshow, 2000).

Table V

Statistical Tests on Final Model M1

Statistical tests on the parsimonious R1+R4 logit. Bootstrap CIs use 500 resamples. Paired-bootstrap p-values for ΔAUC computed against Altman benchmarks on a separate historical Colombian sample where all required Altman inputs are available.

Test	Statistic	p-value / 95% CI	Decision	Interpretation
Bootstrap CI for AUC	AUC = 0.8257	[0.8189, 0.8294]	Precise	Width = 0.011, indicating tight estimation
Bootstrap CI for transition AUC	AUC = 0.6543	[0.6426, 0.6608]	Precise	Width = 0.018, useful precision for transition
Wald joint test	$\chi^2 = 5{,}894$	$p < 0.0001$	Reject H_0	R1 and R4 are jointly highly significant
LR test M1 vs M2	LR = 18.92	$p = 0.000014$	Reject H_0 at large N	But ΔAUC ≈ 0.0000: M1

				preferred on parsimony
Hosmer-Lemeshow	$\chi^2 = 2695$	$p < 0.001$	Reject H_0 (artifact)	Large-N artifact; ECE = 0.019 indicates good calibration
Paired bootstrap vs Altman 1968	ΔAUC = +0.1917	[+0.1699, +0.2145]	Reject H_0	Local significantly better by 19.2 pp
Paired bootstrap vs Altman Z"	ΔAUC = +0.1502	[+0.1295, +0.1703]	Reject H_0	Local significantly better by 15.0 pp

Table VI reports the Murphy (1973) decomposition of the Brier Score into reliability (calibration), resolution (discrimination), and uncertainty (climatology). The decomposition satisfies the identity Brier = Reliability − Resolution + Uncertainty. Excellent reliability (0.0006) indicates that predicted probabilities closely match observed rates within bins; substantial resolution (0.0208) indicates that predictions differ meaningfully from the unconditional rate; and the resulting Brier Skill Score of 0.311 quantifies a substantial improvement over the climatological prior.

Table VI

Brier Score Decomposition (Murphy, 1973)

Decomposition of Brier Score = Reliability − Resolution + Uncertainty. Reliability measures calibration (lower better, ideal = 0). Resolution measures discrimination (higher better). Uncertainty is the climatological term $\bar{P}(1-\bar{P})$. Resolution computed as the residual that satisfies the decomposition identity exactly: Res = Rel + Unc − Brier = 0.000594 + 0.064983 − 0.044799 = 0.020778.

Component	Value	Ideal	Interpretation
Reliability	0.000594	0	Excellent calibration
Resolution	0.020778	higher	Substantial discrimination
Uncertainty	0.064983	fixed	$\bar{P}$ = 6.99% baseline
Brier Score (total)	0.044799	0	Rel − Res + Unc (identity)
Brier Skill Score	0.3106	1	31.1% improvement over climatology

D. Walk-Forward Out-of-Time Validation

Table VII reports walk-forward validation: for each test year t, the model is trained on observations from 2018 through t-1 and evaluated on year t with outcome at t+1. The procedure preserves temporal order and prevents information leakage from future to past. Winsorization thresholds (1st and 99th percentiles) are estimated using only the training sample and applied to the test year, preventing look-ahead bias. The test on 2020 uses financial statements observed during the COVID-19 pandemic to predict accounting distress in the following year (2021). The resulting AUC of 0.84 demonstrates that the model retains predictive discrimination for firms whose statements were observed during the pandemic period; we do not claim the model anticipated distress occurring within 2020 itself.

Table VII

Walk-Forward Out-of-Time Validation (M1: R1+R4)

Walk-forward validation for the parsimonious R1+R4 specification. Train on years through t-1, test on year t with forward-looking outcome at t+1. AUC 95% CIs via 200 bootstrap resamples per test year. BSS = Brier Skill Score; ECE = Expected Calibration Error; MCE = Maximum Calibration Error.

Train ≤	Test	N test	Distress %	AUC	AUC 95% CI	BSS	ECE	MCE
2018	2019	23,183	8.01%	0.7800	[0.768, 0.792]	0.2646	0.0200	0.0439
2019	2020	26,762	6.23%	0.8452	[0.828, 0.850]	0.3124	0.0237	0.0770
2020	2021	26,866	6.85%	0.8327	[0.822, 0.846]	0.3205	0.0192	0.0725
2021	2022	29,134	6.97%	0.8147	[0.803, 0.827]	0.2954	0.0182	0.0636
2022	2023	27,824	6.86%	0.8431	[0.833, 0.855]	0.3125	0.0149	0.0801
Avg	—	—	—	0.8231	—	0.3011	0.0192	0.0674

Out-of-time AUC ranges from 0.78 to 0.85, with an average of 0.82. Brier Skill Score averages 0.30, and average ECE is 0.019.

E. Calibration Analysis and Population Stability

E.1. Decile-Level Calibration

Table VIII reports decile-level calibration for the recommended M1 specification. Firms are sorted into ten equal-sized deciles by predicted probability; we compare average predicted PD against observed distress rate within each decile. The calibration regression of observed distress rate on predicted PD yields slope = 1.1022 (95% CI: [1.0861, 1.1246]), intercept = -0.0071 ([-0.0086, -0.0063]), and $R^2 = 0.9638$, indicating excellent aggregate calibration.

Table VIII

Decile-Level Calibration (M1, Forward-Looking)

Calibration table for the parsimonious R1+R4 logit on the forward-looking sample (N = 151,710). Decile 0 contains firms with the lowest predicted distress probability; decile 9 contains the highest. Lift = decile distress rate / unconditional distress rate (6.99%).

Decile	N	Defaults	PD predicted	Distress observed	Error	Lift
0	15,171	396	0.42%	2.61%	-2.19pp	0.37x
1	15,171	276	0.92%	1.82%	-0.90pp	0.26x
2	15,171	274	1.27%	1.81%	-0.54pp	0.26x
3	15,171	330	1.73%	2.18%	-0.44pp	0.31x
4	15,171	334	2.39%	2.20%	+0.19pp	0.32x
5	15,171	392	3.33%	2.58%	+0.74pp	0.37x
6	15,171	493	4.71%	3.25%	+1.46pp	0.47x
7	15,171	710	7.00%	4.68%	+2.32pp	0.67x
8	15,171	1,123	11.83%	7.40%	+4.42pp	1.06x
9	15,171	6,271	36.27%	41.34%	-5.06pp	5.92x

We disclose one limitation of decile ranking: the lowest deciles are not perfectly monotonic. Decile 0 exhibits an observed distress rate of 2.61%, while deciles 1 and 2 show 1.82% and 1.81%, respectively. The expected ordering reasserts itself from decile 2 onward, with deciles 3-9 strictly increasing. This non-monotonicity in the safest segment likely reflects the score's reliance on a small number of ratios that cannot fine-grain rank firms in the low-risk region. We interpret this as evidence that the score is most useful for identifying high-risk firms rather than for fine-grained ranking among low-risk firms. For high-risk identification, lift in the top decile (5.92x) and in the top two deciles (3.49x) is robust: the top 10% captures 59.2% of distress events, and the top 20% captures 69.8%.

E.2. Population Stability

Table IX reports the Population Stability Index (PSI) of R1, R4, and the predicted probability of distress, computed against a 2018 baseline. PSI values below 0.10 are conventionally interpreted as indicating no material distribution shift; values between 0.10 and 0.25 indicate moderate shift; values above 0.25 indicate major shift. All values across the sample period (2019-2023) remain below 0.10 for both predictors and predicted PD, indicating stable population characteristics.

Table IX

Population Stability Index (vs 2018 Baseline)

Annual PSI for the predictors R1 and R4, and for the predicted probability of distress, computed against a 2018 base distribution. Convention: PSI < 0.10 = stable; 0.10-0.25 = moderate shift; > 0.25 = major shift.

Year	**PSI of R1**	PSI of R4	**PSI of $\hat{P}D$**	**Classification**
2019	0.0102	0.0006	0.0038	✓ Stable
2020	0.0209	0.0095	0.0063	✓ Stable
2021	0.0270	0.0114	0.0280	✓ Stable
2022	0.0620	0.0078	0.0431	✓ Stable

2023	0.0434	0.0216	0.0499	✓ Stable
Avg	0.0327	0.0102	0.0262	All years stable

Note: PSI of R4 is reported separately. The stability of $\hat{P}D$ additionally confirms joint stability of (R1, R4).

F. Pooled Logit with Year and Sector Fixed Effects

Table X reports the hazard-style pooled logit with year and sector fixed effects (M3). We include year dummies for 2019-2023 and sector dummies for the top 30 two-digit CIIU divisions. Standard errors are clustered by NIT.

Table X

Hazard-Style Pooled Logit with Year and Sector Fixed Effects

Logit estimates with year and sector fixed effects on the forward-looking sample. Column 1 reports R1+R4 baseline (no FE); Column 2 adds year and sector fixed effects. Standard errors clustered by NIT. Year and sector fixed effect estimates suppressed for brevity.

Variable	M1 (no FE)	M3 (Year + Sector FE)
const	-1.2011*** | (-54.45)	-1.5618*** (-19.93)
R1	-6.7120*** | (-43.28)	-6.5782*** (-44.11)
R4	-3.4259*** | (-56.48)	-3.4386*** (-57.35)
Year FE	No	Yes
Sector FE	No	Yes (top 30)
N	151,710	151,710
AUC	0.8257	0.8340
Brier	0.0448	0.0444

The fixed-effects specification yields modestly higher AUC (0.834 vs 0.826), an improvement of 0.83 percentage points. R1 and R4 coefficients are economically and statistically similar across specifications, with sign and approximate magnitude preserved. The result confirms that the bulk of forward-looking discriminative power resides in the two ratios themselves; fixed effects provide a modest improvement consistent with sectoral and macroeconomic heterogeneity.

G. Apples-to-Apples Comparison with Altman Benchmarks

An accurate comparison with Altman benchmarks requires that all models be evaluated on the same sample with the same outcome definitions. Two issues arise. First, Altman scores require Working Capital and Retained Earnings, which are not reported in the post-IFRS Supersociedades extract for 2018-2024. Second, naive comparison of Altman benchmarks computed on the full sample with the local model computed on the forward-looking sample violates the apples-to-apples principle.

We address both issues by computing all benchmarks on a separate historical Colombian panel (2006-2018, N = 30,057 forward-looking) where Working Capital, Retained Earnings, and all other Altman inputs are available. We emphasize that this comparison sample is distinct from the main 2018-2024 panel; differences in absolute AUC values between Tables VII and XI partly reflect this sample difference. The higher local AUC on the pre-IFRS panel (0.845 vs 0.826 on the main 2018-2024 panel) is consistent with the asset-revaluation pattern documented in Section VI.B: post-IFRS accounting introduces upward revaluations that inflate book equity in a subset of distressed firms (the Coltejer pattern), partially attenuating R4's discriminative power; pre-IFRS Colombian accounting under Decree 2649 used more conservative historical-cost rules that preserved R4 as a sharper distress signal. Altman scores are inverted to match the AUC convention for distress prediction (higher score = riskier firm). We test the difference in AUC against each Altman benchmark via paired bootstrap with 500 resamples.

Table XI

Apples-to-Apples Comparison: Altman vs Local, Historical Panel

Comparison on the separate historical Colombian panel 2006-2018, where all Altman inputs (Working Capital, Retained Earnings, and other components) are available. Altman scores inverted to match distress-AUC convention. ΔAUC tested via paired bootstrap (500 resamples).

Model	AUC	ΔAUC vs Local	95% CI of ΔAUC	Bootstrap p-value
Local R1+R4 logit (forward)	0.8452	— (reference)	—	—
Altman 1968 (oriented)	0.6534	-0.1917	[-0.2145, -0.1699]	≈ 0.002
Altman Z" 1995 (oriented)	0.6949	-0.1502	[-0.1703, -0.1295]	≈ 0.002

Table XI demonstrates two results. First, the Altman benchmarks, when correctly oriented, achieve AUC of 0.65 (1968) and 0.69 (Z") on the historical Colombian forward-looking sample. These are not 'worse than random,' contrary to a naive comparison that fails to invert the score. Second, the local R1+R4 logit improves over the best Altman benchmark by 15.0 percentage points and over Altman 1968 by 19.2 percentage points, with paired-bootstrap p-values below 0.005 in both cases. Local calibration provides a real and statistically significant—but moderate—improvement over imported Altman scores when both are evaluated honestly on the same sample.

To address the potential objection that this comparison favors the local model because Altman coefficients were imported rather than locally estimated, we conduct an additional apples-to-apples comparison against a recalibrated Altman benchmark. Specifically, we estimate a reduced Altman specification (X3 = EBIT/TA, X4 = Book Equity/Total Liabilities, X5 = Sales/TA) by logistic regression on Colombian data 2018-2022 with the same cluster-robust standard errors by NIT, and compare it to M1 on the full forward-looking panel. On 161,253 firm-year observations, M1 achieves AUC = 0.8208 versus 0.7872 for the recalibrated reduced Altman and 0.7428 for the imported reduced Altman, with paired-bootstrap differences of +3.36 percentage points (95% CI: [+2.93, +3.77]) and +7.78 percentage points (95% CI: [+7.44, +8.13]) respectively. M1 wins with probability 1.000 against both versions in 200 bootstrap resamples. The recalibration of the Altman benchmark reveals that on Colombian data only X3 (EBIT/TA) remains statistically significant; X4 and X5 lose discriminative power. This reinforces the parsimony argument: two well-chosen ratios outperform a locally-optimized three-ratio specification.

H. Sectoral Robustness

We re-estimate the parsimonious M1 specification separately within each of the fifteen largest two-digit CIIU sectors. Sector-level forward-looking AUC ranges from 0.694 (CIIU K64, monetary intermediation) to 0.881 (CIIU 47, retail trade). Heterogeneity is real but manageable. The model performs strongest in commerce, agriculture, and food (AUC > 0.85), and weakest in financial-services sectors (K64 = 0.694, K66 = 0.724); see structural interpretation below. Coefficient signs are preserved everywhere (both negative as expected), with magnitudes ranging from 2.9 to 8.2 on R1 and from 1.8 to 4.9 on R4. We recommend users of the score consider sector-specific recalibration, particularly in financial services.

We provide a structural interpretation for the weak performance in financial sectors (CIIU K64 monetary intermediation and K66 financial auxiliaries). A controlled subsample comparison shows aggregate financial-sector AUC = 0.7085 (95% bootstrap CI: [0.682, 0.730]) versus 0.8107 ([0.792, 0.832]) for a random control of the same size from the real sector. The difference is structural, not driven by sample size. The structural driver is visible in the predictor distribution: the median R4 in K64+K66 is 0.86, compared to 0.50 in the real sector. The entities classified in K64/K66 in the Supersociedades panel are predominantly holding companies, investment vehicles, and fiduciaries rather than deposit-taking banks subject to prudential capital regulation. Their structurally high book equity saturates the capitalization ratio near its upper bound and removes its discriminative power, leaving the model effectively reliant on R1 alone. The formal model is therefore best understood as out of scope for true financial intermediaries, where sector-specific specifications incorporating prudential ratios (loan quality, coverage, liquidity) would be appropriate.

I. Hazard Model as Robustness Check

To verify that the choice of a pooled logit with fixed effects does not mask particular sensitivities of a duration specification, we estimate a Cox proportional hazards model with time-varying covariates R1 and R4 on the same forward-looking panel. The specification defines each NIT (firm) as the subject, entry into contemporaneous distress as the event, and right-censoring for firms that exit the panel without entering distress. The model is fit on 38,090 unique firms with 173,135 firm-year intervals and 9,350 observed events. The Cox subsample exceeds the main forward-looking panel of 37,653 firms (Table I) because the duration analysis includes 437 additional firms with at least one observed firm-year that contribute valid right-censored intervals to the risk set, even though they lack the consecutive (t, t+1) observation required for the main panel.

Estimated coefficients are $\beta\_R1 = -11.22$ and $\beta\_R4 = -0.35$, both with $p < 1e-6$ and signs consistent with the M1 logit. The coefficients are stable to (i) the absence of regularization (penalizer = 0 reproduces the same estimates), (ii) stratification by CIIU industry division ($\Delta\beta < 3\%$ in both coefficients), and (iii) out-of-time validation (training on 2018-2021 and evaluating on 2022-2023 reproduces the coefficients and yields AUC = 0.870). Adding R1·log(t) and R4·log(t) interactions reveals statistically significant non-proportionality in both covariates ($p < 1e-6$), interpretable as an intensification of R1 effects with firm tenure in the panel and an attenuation of R4 effects.

The Cox C-statistic on Distress_{t+1} is 0.876, six percentage points above the M1 logit (0.810) on the same sub-panel. This gap should not be interpreted as structural superiority of the duration specification. An apples-to-apples comparison with a logit re-estimated on the same sub-panel yields AUC = 0.878 -- essentially identical to the Cox -- indicating that the apparent gain reflects local recalibration of coefficients rather than properties of the baseline hazard. The official M1 logit with coefficients estimated on the full 2018-2024 sample remains the recommended operational specification; the Cox confirms the direction and dominance pattern of the logit and is reported here only as a duration-based robustness check.

A useful conceptual interpretation of the Cox results: R1 (operating profitability) operates primarily as a predictor of the moment of entry into the distress state, while R4 (capitalization) operates primarily as a predictor of persistence in that state. The forward-looking logit captures both channels jointly because its outcome DISTRESS_{t+1} is a state indicator that conflates entry and persistence; the Cox time-varying specification decomposes them by modeling only the entry hazard and filtering already-distressed firms from the risk set. R4 retains a statistically significant but quantitatively smaller role in the Cox (LRT = 205, $p < 1e-6$), consistent with the interpretation that its main predictive value lies in the persistence channel that the logit captures but the Cox does not. The two specifications are complementary, not competing.

J. External Cross-Reference Against Legal Insolvency Events

Sections III-V calibrate and validate the score on accounting distress, defined by accounting criteria (negative equity, liabilities exceeding assets, or severe operating losses). The model is not estimated on legal events of insolvency. This section provides an external, post-hoc cross-reference between predicted PD values and the Supersociedades public registry of legal insolvency proceedings opened between 2022 and 2025.

Procedure. The Supersociedades registry distinguishes between two principal proceeding types under Law 1116 of 2006: reorganization or restructuring agreements (reorganizacion) and judicial liquidation (liquidacion judicial); the registry also includes a small set of other proceedings (acuerdos de adjudicacion, validation of out-of-court agreements). We match firms in the financial panel to the insolvency registry by normalized NIT and identify the year of opening of each legal proceeding. For each matched firm, we compute the predicted PD using R1 and R4 measured one year before the legal event, applying the operational formula in equations (10)-(11) and the same winsorization thresholds as the main analysis.

Findings. Table XII reports the resulting PD distribution by proceeding type. The median predicted PD one year before opening is 43.7% for liquidations (mean 51.1%); 80.5% of liquidations show PD >= 10% and 60.3% show PD >= 25% one year prior. Reorganizations show weaker but still informative concentration: median 10.7%, with 52.9% above the 10% threshold.

Table XII

Predicted Distress Probability One Year Before Legal Insolvency Events

Distribution of predicted PD at year t-1 by type of insolvency proceeding registered between 2022 and 2025. Sample restricted to firms with NIT matched to the financial panel and financial data available at t-1. PD computed using equations (10) and (11) with the same winsorization as the main analysis.

Proceeding type	**N**	**Mean PD t-1**	**Median PD t-1**	**PD ≥ 10%**	**PD ≥ 25%**	**PD ≥ 50%**
Judicial	655	51.1%	43.7%	80.5%	60.3%	47.5%

liquidation						
Reorganization / restructuring	2,883	21.1%	10.7%	52.9%	22.1%	10.5%
Other proceedings (Ley 1116)	37	11.8%	9.0%	—	13.5%	—
Total legal insolvency	**3,575**	**26.5%**	**12.4%**	**57.9%**	**29.0%**	**16.7%**

Interpretive cautions. This evidence is asymmetric. We observe the PD distribution among firms that subsequently experienced a registered legal event, which is informative about whether high PD precedes legal insolvency; we do not observe the PD distribution at t-1 in the broader population of firms that did not experience legal events. The cross-reference therefore cannot quantify the false-positive rate at any given PD threshold in the population, nor the operational tradeoff a credit officer would face when using a threshold rule. Stated differently: we can say that a high PD precedes legal liquidation in a substantial share of cases; we cannot say from this analysis alone how many firms with high PD do not enter legal insolvency. The asymmetry is inherent in cross-referencing against a registry of realized events; addressing it would require a population-level outcome variable for legal events that the available data do not provide. We additionally note that the registry covers 2022-2025, which overlaps temporally with the main estimation period; the cross-reference therefore does not constitute genuine out-of-sample evaluation in the strict temporal sense.

Implication. The concentration of high predicted PD among firms that subsequently entered judicial liquidation is consistent with the interpretation that the accounting-based score also responds to the deteriorating fundamentals that precede legal events. We do not claim that the score is calibrated for, or validated as a predictor of, legal insolvency. Sections V.B-V.G calibrate and validate the score on accounting distress.

VI. Discussion

A. Implications for Practitioners

Our results have implications for several practitioner audiences. For credit-risk officers, the forward-looking AUC of 0.82 and top-decile lift of 5.9 suggest that the score can guide monitoring resources, although the transition AUC of 0.65 tempers expectations of identifying brand-new failures. For investment-banking due diligence, the parsimony of two ratios facilitates deployment on standard financial statement data. For regulatory supervision, the stable walk-forward performance, including in 2020, supports the use of the score as a complementary monitoring tool.

B. Limitations

Sample selection. The analytical panel comprises formal corporates that exceed Supersociedades' statutory reporting thresholds. Small firms, informal firms, and entities below threshold are not in the sample. The estimated coefficients and the score's discriminative power apply to firms with this size and formalization profile and do not generalize to the broader Colombian corporate universe.

Sectoral scope. Sectoral robustness (Section V.H) documents that the score's discriminative power is materially weaker in CIIU divisions K64 (monetary intermediation, AUC = 0.694) and K66 (financial auxiliaries, AUC = 0.724), where the median capitalization ratio (R4 = 0.86) is substantially higher than in the real sector (R4 = 0.50) and saturates the predictor near its upper bound. The entities classified in these divisions in the Supersociedades panel are predominantly holding companies, investment vehicles, and fiduciaries rather than deposit-taking banks subject to prudential capital regulation. We do not recommend the score for credit-risk assessment of entities with financial-sector balance-sheet structures; sector-specific specifications incorporating prudential ratios (loan quality, coverage, liquidity) would be appropriate for those entities and are outside the present analysis.

Temporal coverage. The estimation period (2018-2024) captures the COVID-19 shock of 2020-2021 but does not span a complete Colombian credit cycle. The Colombian macro-credit cycle includes periods of more pronounced deterioration (1998-2002 banking crisis, 2008-2009 global financial crisis) that are entirely outside the post-IFRS estimation window. Whether the parsimonious specification's stability under the COVID shock generalizes to more severe credit-cycle episodes is an open question that the available data cannot directly address.

Outcome definition. The primary outcomes are accounting-based. While Section V.J provides an external cross-reference against legal insolvency events, the score is calibrated on accounting distress, and the legal-event evidence is asymmetric in the sense made precise in Section V.J. We do not claim that the score is a validated predictor of legal insolvency.

Survivorship bias. The forward-looking panel conditions on firms observed in both year t and year t+1. Section III.C documents that attrited firms have year-t distress rates approximately 3.7

times those of persisting firms; the direction of the resulting bias in AUC estimates is theoretically ambiguous, depending on whether attrited distress cases are harder or easier to classify than persisting distress cases. The sensitivity analysis reported in the introduction shows that the forward-looking AUC is robust under plausible imputations of attrited outcomes (a 0.09 percentage-point reduction under the assumption that all attriters are healthy, a 0.68 percentage-point increase under selective imputation; only the empirically implausible assumption that all attrited firms are distressed produces a material 7.95-percentage-point reduction). A more standard panel-econometric treatment would be inverse-probability weighting (IPW) based on a model of P(attrition | observed characteristics); we report sensitivity scenarios rather than IPW because the variables that plausibly drive selection (firm size below threshold, mergers, ownership transfers) are not all observable in the available extract. Implementing IPW with the observable subset of predictors would be a natural extension.

Calibration in the low-risk region. Aggregate calibration of the parsimonious specification is good (slope = 1.10, intercept = -0.01, R2 = 0.96 on the decile-level calibration regression), but the lowest-risk deciles are not perfectly monotonic: decile 0 has an observed distress rate of 2.61%, while deciles 1 and 2 have 1.82% and 1.81%, respectively. The score is most useful for identifying firms in the high-risk region (top-decile lift = 5.92x; top 10% captures 59.2% of distress events) and is not appropriate for fine-grained ranking among firms in the safest deciles. Practitioners should treat the score as a triage tool for monitoring resources rather than as a full ordering of firms by credit risk.

Asset-revaluation pattern. Section V documents 1,227 firm-years where high reported capitalization ($R4 > 0.85$) coexists with operating losses ($R1 < -0.05$); these firms experience distress in t+1 at a rate of 23.3%, versus the score's mean prediction of 11.1%. The pattern is consistent with post-IFRS upward revaluations of fixed assets or financial investments inflating book equity in firms whose operations are deteriorating. Users of the score should inspect the source of recent equity growth before relying on R4 for firms exhibiting this profile.

Performance metrics reported. Performance is reported via ROC-AUC with bootstrap CIs, Brier decomposition (Murphy, 1973), ECE/MCE, decile-level calibration, and lift. Standard operational metrics for binary credit-risk classifiers - precision-recall AUC, precision and recall at named PD thresholds (10%, 25%, 50%), and confusion matrices at recommended operating points - are not reported in this draft and would be a natural extension; they are computable on the existing data and would provide additional detail for the imbalanced-outcome setting (base rate approximately 7%).

Cross-country external validity. The framework is portable; the Colombian point estimates are not. Cross-country replication for Mexico, Brazil, Chile, Peru, and Argentina would test the generalizability of both the three-estimand gap and the parsimonious specification to other emerging markets.

VII. Conclusion

Accounting-based distress prediction is a setting in which the predictor set and the outcome share an accounting identity, generating a form of target leakage that is structural rather than incidental. We propose and apply a three-estimand framework - contemporaneous classification, forward-looking prediction, and transition prediction - that decomposes the leakage contribution into a measurable gap across the three. Applied to 192,458 firm-year observations from the Colombian regulator Supersociedades for 2018-2024, the three estimands yield AUC values of 0.955, 0.826 and 0.654, with the 13-30 percentage-point gap quantifying the magnitude of the definitional dependence component.

A parsimonious specification combining operating profitability and capitalization (R1 = EBIT/Total Assets, R4 = Equity/Total Assets) captures the predictive content; a candidate third predictor adds essentially no forward-looking discrimination. The specification is stable across walk-forward out-of-time tests including the COVID-19 year, improves AUC by 15-19 percentage points over correctly-oriented Altman benchmarks on a separate historical panel where all Altman inputs are available, and produces a 3.4-percentage-point AUC advantage over a locally re-estimated reduced Altman specification - consistent with the interpretation that the bias-variance tradeoff in this setting favors a small number of theoretically grounded predictors over richer specifications.

The methodological contribution is portable to settings beyond Colombian private firms and beyond distress prediction: any predictive problem in which the predictor set and the outcome share an accounting identity admits the same three-estimand decomposition. The Colombian point estimates illustrate the framework rather than constituting its main claim. The score is not recommended for entities with financial-sector balance-sheet structures (CIIU K64, K66).

Several extensions are natural. Cross-country replication across Latin America would test whether the three-estimand gap is similarly large in other emerging-market accounting regimes. Integration of legal insolvency events as a properly estimated outcome - rather than as a post-hoc cross-reference - would require access to a population-level outcome variable that the current data do not provide. Inverse-probability weighting for the panel-attrition correction, the addition of standard operational metrics (PR-AUC, precision and recall at named thresholds, confusion matrices), machine-learning extensions for the low-risk region where the parsimonious logit exhibits non-monotonic ranking, and sector-specific specifications for financial intermediaries are all promising directions.

Appendix A. Operational Scoring Formulas and Worked Examples

We provide explicit scoring formulas for the recommended specification and a complementary transition specification. All formulas yield a probability of accounting distress in the following year. Coefficients are estimated on the full forward-looking sample (N = 151,710) for M1 and on the transition sample (N = 142,214) for the transition specification.

A.1. Recommended forecasting score (M1: R1 + R4)

For a firm with operating profitability R1 = EBIT_t / TA_t and capitalization R4 = Equity_t / TA_t, the predicted probability of accounting distress at t+1 is given by Equations (10) and (11) reproduced here for convenience:

$$z = -1.2011 + (-6.7120)\cdot R1 + (-3.4259)\cdot R4 \qquad (A1)$$

$$P(DISTRESS_{t+1} = 1) = 1 / (1 + \exp(-z)) \qquad \text{(A2)}$$

A.2. Transition specification (firms not currently distressed)

$$z = -2.4172 + (-3.9491)\cdot R1 + (-1.4902)\cdot R4 \qquad (A3)$$

This specification is appropriate when applying the score to firms not currently in financial distress (Outcome C). It exhibits smaller magnitude coefficients, reflecting that genuine new-deterioration is harder to predict than persistent distress. The transition AUC is 0.6543 (95% CI: [0.6426, 0.6608]).

A.3. Worked examples

Example 1 — Below-average risk firm. Consider a firm with EBIT/TA = 0.05 (5% operating profitability) and Equity/TA = 0.40 (40% capitalization). Applying the recommended formula:

$z = -1.2011 + (-6.7120)\cdot(0.05) + (-3.4259)\cdot(0.40) = -2.9071$

$P(DISTRESS) = 1 / (1 + \exp(2.9071)) = 0.0518 = 5.18\%$

This firm is below the unconditional distress rate of 6.99%, indicating below-average risk.

Example 2 — High-risk firm. Consider a firm with EBIT/TA = -0.02 (small operating loss) and Equity/TA = 0.10 (highly leveraged). Applying the same formula:

$z = -1.2011 + (-6.7120)\cdot(-0.02) + (-3.4259)\cdot(0.10) = -1.4095$

$$P(DISTRESS) = 1 / (1 + \exp(1.4095)) = 0.1963 = 19.63\%$$

This firm is approximately 2.8 times the unconditional rate, placing it in the upper deciles of predicted risk. It would be flagged for credit-risk monitoring under most operational thresholds.

Declarations

Data Availability

The data used in this study are obtained from the Superintendencia de Sociedades (Supersociedades) Estados Financieros database, publicly available at https://www.supersociedades.gov.co/portafolio-de-servicios/sirem. The version used corresponds to firm-year filings for fiscal years 2018-2024, downloaded in the form of consolidated annual statements per CIIU classification. Replication code (Python, including data preparation, model estimation, bootstrap procedures, and table generation) and a snapshot of the cleaned analytical panel will be deposited in a public repository (Open Science Framework or Zenodo) at the time of paper acceptance, with persistent DOI assigned for citation. The repository will include: (i) the full Python pipeline; (ii) the analytical panel (deidentified at the firm level); (iii) intermediate output files; (iv) figures and tables in source form; and (v) version specification of all software dependencies.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Conflicts of Interest

The author declares no conflicts of interest. The author works in investment banking and corporate valuation in Colombia; no employer or client of the author has reviewed, sponsored, or influenced the content of this paper.

Author Contributions

Jiménez Triviño is the sole author and is responsible for all aspects of the research, including conceptualization, methodology, data curation, formal analysis, software development, validation, visualization, writing of the original draft, and writing of the revised version.

Ethics Statement

This research uses publicly available regulatory financial statement data and does not involve human subjects, personal information, or experimental interventions. No ethics approval was required.

References

Adams, J., D. Hayunga, S. Mansi, D. Reeb, and V. Verardi, 2019, Identifying and treating outliers in finance, Financial Management 48, 345-384.

Altman, E. I., 1968, Financial ratios, discriminant analysis and the prediction of corporate bankruptcy, Journal of Finance 23, 589-609.

Altman, E. I., 1995, Emerging market corporate bonds: A scoring system, in The Future of Emerging Market Flows, J. P. Morgan & Salomon Brothers, 23-30.

Bai, Y., and C. J. Green, 2020, Country and industry factors in tests of capital asset pricing models for partially integrated emerging markets, Economic Modelling 92, 180-194.

Beaver, W. H., 1966, Financial ratios as predictors of failure, Journal of Accounting Research 4, 71-111.

Caicedo-Cerezo, E., 2006, Análisis de discriminante aplicado al modelo de predicción de quiebra para empresas colombianas, Cuadernos de Administración 19, 31-66.

Campbell, J. Y., J. Hilscher, and J. Szilagyi, 2008, In search of distress risk, Journal of Finance 63, 2899-2939.

DeLong, E. R., D. M. DeLong, and D. L. Clarke-Pearson, 1988, Comparing the areas under two or more correlated receiver operating characteristic curves: A nonparametric approach, Biometrics 44, 837-845.

Hosmer, D. W., and S. Lemeshow, 2000, Applied Logistic Regression, 2nd ed. (Wiley).

Hyndman, R. J., and G. Athanasopoulos, 2018, Forecasting: Principles and Practice, 2nd ed. (OTexts).

Kaufman, S., S. Rosset, and C. Perlich, 2012, Leakage in data mining: Formulation, detection, and avoidance, ACM Transactions on Knowledge Discovery from Data 6, Article 15.

Murphy, A. H., 1973, A new vector partition of the probability score, Journal of Applied Meteorology 12, 595-600.

Ohlson, J. A., 1980, Financial ratios and the probabilistic prediction of bankruptcy, Journal of Accounting Research 18, 109-131.

Pereiro, L. E., 2001, The valuation of closely-held companies in Latin America, Emerging Markets Review 2, 330-370.

Pereiro, L. E., 2010, The beta dilemma in emerging markets, Journal of Applied Corporate Finance 22, 110-122.

Sarmiento-Sabogal, J., and M. Sadeghi, 2014, Unlevered betas and the cost of equity capital: An empirical approach, North American Journal of Economics and Finance 30, 90-105.

Sarmiento-Sabogal, J., and M. Sadeghi, 2015, Estimating the cost of equity for private firms using accounting fundamentals, Applied Economics 47, 288-301.

Sarmiento-Sabogal, J., M. Sadeghi, J. Sandoval, and E. Cayón, 2021, The application of proxy methods for estimating the cost of equity for unlisted companies: evidence from listed firms, Review of Quantitative Finance and Accounting 57, 1009-1031.

Shumway, T., 2001, Forecasting bankruptcy more accurately: A simple hazard model, Journal of Business 74, 101-124.